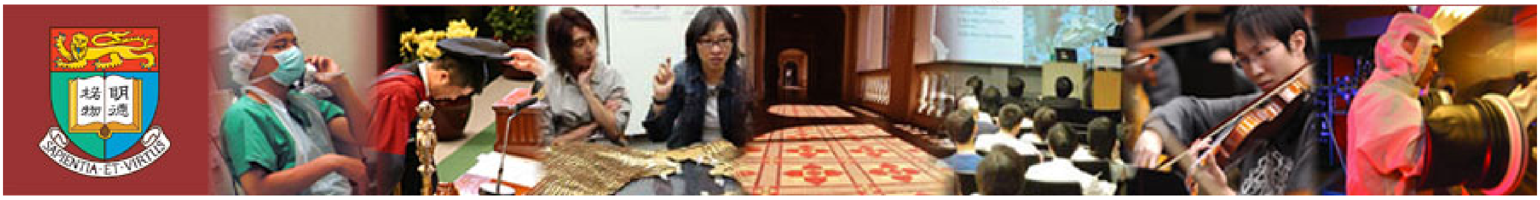


Title	Using prosody in Cantonese-speaking children with severe dysarthria: caregiver and speaker variables
Other Contributor(s)	University of Hong Kong.
Author(s)	Lai, Fan-chun, Jenny
Citation	
Issued Date	2007
URL	http://hdl.handle.net/10722/55487
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Using prosody in Cantonese-speaking children with severe dysarthria:

Caregiver and speaker variables

Lai, Fan Chun Jenny

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2007.

ABSTRACT

This study aimed to investigate whether children with severe dysarthria are able to use prosody (pitch and duration) to communicate with their caregivers. Four children with severe dysarthria resulting from cerebral palsy and their caregivers participated in two tasks. The children were asked to produce the vowel /a/ at 5 pitch levels and 5 durations; the caregivers were asked to identify the target the children produced. Productions were also analyzed acoustically. Both caregiver accuracy and acoustic productions varied across subjects. In general, children were able to produce some pitch and duration distinctions. Poor correspondence between accuracy and acoustic measures suggested that caregivers might use other cues in perception. This study suggested that children with severe dysarthria have the ability to use prosody to signal contrasts, which might be developed as a means to interact with their caregivers or with augmentative and alternative communication devices.

INTRODUCTION

Dysarthria is characterized by disturbances in strength, speed, range, steadiness, tone or accuracy of movements resulting from damage to the central or peripheral nervous system, which affect control of the respiratory, phonatory, resonatory, articulatory, and prosodic aspects of speech production (Darley, Aronson & Brown, 1975; Duffy, 2005). Most studies that have described the characteristics of dysarthria resulting from cerebral palsy have been limited to segmental aspects of speech (Ansel & Kent, 1992; Hardy, 1983; Platt, Andrews, Young & Quinn, 1980; Platt, Andrews & Howie, 1980). Few studies have investigated the prosodic aspects of dysarthric speech. Although increased attention is being given to the role of prosody in dysarthric speech, much remains to be explored (Leuschel & Docherty, 1996; Yorkston, Beukelman & Bell, 1988).

It has been suggested that speakers with significantly reduced ability to control segmental aspects of speech may be able to make use of prosodic cues to deliver their intentions (Le Dorze, Ouellet, & Ryalls, 1994; Patel, 2002b; Vance, 1994). Patel (2002a) found that adults with severe dysarthria resulting from cerebral palsy were able to consistently control sustained production of the vowel /a/ at three distinctive durations and two distinctive pitches. Patel (2002b) also found that dysarthric speakers can use prosodic control to signal question-statement contrasts, and listeners could accurately classify the contrasts. Patel (2004) reported that dysarthric speakers were able to use increased fundamental frequency, intensity and duration to signal stressed words at four word positions in a similar way to nonimpaired healthy controls. The dysarthric group and a non-dysarthric control group did not differ in the extent or range of fundamental frequency to mark stress (Patel, 2004).

These findings have important clinical implications as individuals with severe dysarthria can be trained to use their prosodic control capabilities to convey communication intentions or use them as an input channel for augmentative and alternative communication devices (Patel, 2002a; Patel, 2002b). Although the use of augmentative and alternative communication devices (AAC) is common in children with severe dysarthria, it remains the goal for the children to achieve some sort of oral speech (Love, 2000). Yorkston, Beukelman & Bell (1988) agreed that the use of speech should not be given up once an AAC device is in place and even severely limited speech has significant communicative function.

While there is evidence to suggest that English speakers with severe dysarthria are able to use prosody meaningfully, not much is known about the use of prosody by Cantonese speakers with dysarthria. Cantonese differs from English in that it is a tonal language i.e., tones carry lexical meaning. There are six contrastive tones in Cantonese, namely high level (55), high rising (35), mid level (33), low falling (21), low rising (23) and low level (22). Since lexical tone is contrasted by differences in fundamental frequency (Fok, 1974; Vance, 1976), it is possible that Cantonese listeners are able to interpret variations in fundamental frequency with even more accuracy than listeners from non-tonal languages, such as English. Francis & Ciocca (2003) studied whether experience with a tonal language affected listeners' sensitivity to small differences in fundamental frequency. They found that monolingual English-speaking listeners and native Cantonese-speaking listeners performed similarly in tasks that involved discrimination of speech sounds differing in fundamental frequency. However, Cantonese listeners were less sensitive than English listeners to pitch differences for nonspeech sounds. It is important to study if Cantonese dysarthric speakers can make use of prosody and if Cantonese communication partners can interpret these prosodic cues. In order to find out the answer, the present study targeted at Cantonese-speaking dysarthric speakers and listeners.

Patel & Salata (2006) studied the communication between children with severe dysarthria and their caregivers using prosody through an interactive computer game. Five children with severe dysarthria were asked to produce the vowel /a/ at three distinct pitch levels (high, medium, low), three distinct durations (long, medium, short) and nine combinations of pitch and duration parameters (low and short, low and medium, etc.) in three games respectively. They found that the accuracy for each game varied across caregivers. In addition, Patel & Salata (2006) reported that the caregivers' accuracy for pitch game and duration game were higher than combined game. An extension of Patel & Salata (2006)'s study was conducted in the present study with Cantonese speakers. Two additional pitch levels were added, rising and falling, to correspond with the tone patterns of Cantonese. This resulted in a total of five pitch levels: high, mid, low, rising and falling. Since Patel (2002a) found that severe dysarthric speakers can produce three distinctive durations, it was considered worthwhile to find out if they can produce more distinctive durations. Thus five durations (shortest, short, mid, long, and longest) were targeted in this study. In addition, speakers were asked to produce 11 combined duration and pitch levels (longest duration and low pitch, shortest duration and high pitch, etc.). Due to page limitation, the section on combined game was not presented in this paper.

In addition to caregivers' perception, the children's productions were analyzed acoustically. The analysis included measuring mean fundamental frequency for each pitch level and mean duration for each duration target for each subject.

Furthermore, the present study aimed to compare acoustic results from perceptual results. Studies on normal Cantonese speakers have showed that fundamental frequency (F_0) is the major acoustic cue for pitch perception (Bauer & Benedict, 1997; Fok, 1974). The acoustic correlate for perception of duration is the measured duration of production. Acoustic cues are significant for perception of pitch and duration in normal Cantonese speakers. However, little is known about the relationships between acoustic and perceptual measures in speakers with dysarthria. Thus it is meaningful to find out how and the extent to which acoustic measures corresponds to perceptual measures in speakers with dysarthria.

To conclude, the first research question was to find out caregiver accuracy at interpreting their children's productions of pitch and duration parameters. The second research question was to identify whether Cantonese-speaking children with dysarthria can produce distinctive pitch levels and durations using acoustic analysis. The third research question was to find out the relationship between the perceptual and acoustic findings in speakers with dysarthria.

It was predicted that Cantonese-speaking listeners would be able to perceive prosodic levels at least as well as English-speaking listeners in Patel & Salata (2006). This was supported by the findings of Francis & Ciocca (2003) that English and Cantonese listeners are equally sensitive to differences in the fundamental frequency of speech sounds.

METHOD

Participants

Four male children, ranging in age from 7 to 13 years (mean age of 10.4 years) with severe dysarthria due to cerebral palsy were recruited as speakers from physically handicapped schools. Their caregivers were recruited as listeners. The criteria for speakers were that (a) they had dysarthria as the primary speech diagnosis, (b) they were not using speech as their main means of communication, and that they were using augmentative communication, (c) they had receptive language and cognitive skills adequate for completing the experimental task, and (d) they had adequate hearing ability at conversational speech and adequate visual acuity for looking at a computer screen. These criteria were met according to the referring at each school.

Materials and procedures

The materials and procedures used in the present study were based on those reported by Patel & Salata (2006). A modified “Wizard of Oz” computer game was used. The child was instructed to produce the vowel /a/ at a given level while looking at the computer screen to move a particular figure. The caregiver was seated opposite the child facing another computer and selected the figure he or she believed her child intended to produce. Visual feedback was given to both sides for each trial. Appendix A shows the displays of the child and caregiver interfaces for the pitch and duration games. Using the pitch game as an example, each fish represented one pitch level. For the child’s interface, the location of the worm indicated the pitch target that was being requested. The caregiver’s interface did not show the worm (target). The caregiver was asked to identify each vocalization as high, mid, low, rising or falling by clicking on the fish that he or she predicted the child intended to produce. The chosen fish would swim and get the worm if the selection was correct, otherwise it failed.

The order of administration of the pitch and duration games was counterbalanced among participants. At the start of each game, the experimenter introduced the target that each figure represented and gave examples for each one. Then the child was asked to practice vocalization at each target three times. The participants were given three trials to practice for each game with specific verbal feedback before data collection started. Approximately five seconds was allowed between each trial. If a speaker produced a vocalization and then spontaneously corrected himself or herself, the self-correction was accepted as the actual trial; only one reattempt was allowed. A rest period of five minutes was provided between each game.

Five vocalizations were requested at each target for each game in randomized order. Speech recordings were made using Toplux M-750 digital audio recorder. SONY LAV 05 microphone with lapel attachment was placed at 10 cm - 15 cm from the mouth of the speaker. The mouth-to-mouth distance was monitored throughout the session. The experiment was conducted in a silent room at the school that the child is attending or at his apartment.

Acoustic analysis

Acoustic analysis was performed using the PRAAT System (Boersma & Weenink, 2000). Each vocalization collected at the pitch game was measured at five consecutive and equally spaced time points positions of the vocalic segments (initial, 25%, 50%, 75%, final). Mean F_0 and standard deviation were calculated at each pitch level for each speaker. For vocalizations collected at the duration game, the beginning and end of waveforms were manually marked. Mean duration and standard deviation were calculated at each duration for each speaker. A sample of the acoustic signal was shown in Appendix B.

The experimenter re-analyzed 10% of the samples which were randomly selected from the pitch and duration games; intra-rater reliability calculated by Pearson's correlation coefficient was 0.998. A second examiner performed the analysis for 10% of the samples which were randomly selected from the pitch and duration games; inter-rater reliability calculated by Pearson's correlation coefficient was 0.990.

RESULTS

The results are presented below in the form of caregiver accuracy, caregiver errors and acoustic data, for pitch targets followed by duration targets.

Pitch levels

Figure 1 shows caregiver accuracy for the five target pitch levels for each subject. As can be seen, caregiver accuracy varied across subjects. Mid pitch was identified with 40% accuracy for S2, S3 and S4, and with 100% accuracy for S1. For S3, falling pitch was identified with the highest accuracy. S2 and S4 had accuracy at or below 40% for all pitch levels. In addition, it is noted that S1, S2 and S4 attained better than chance level accuracies (20%) for three of the five targets; S3 for two of the five targets. These findings are discussed further in the context of the error data and acoustic data presented below, case-by-case.

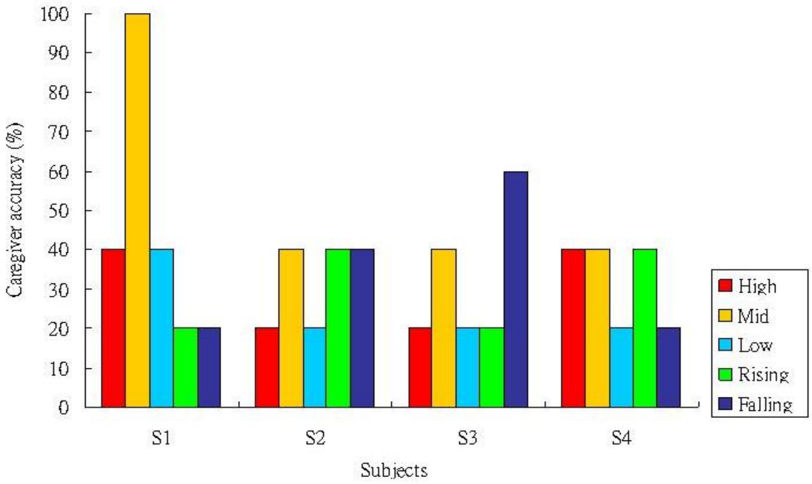


Figure 1. Caregiver accuracy for pitch targets.

<u>Subject 1</u>

Subject 1 achieved high, mid and low pitches with higher than chance accuracies, and mid pitch with 100% accuracy. Although high and mid pitches achieved above-chance accuracies, the acoustic data was surprising. Figure 2 shows the mean F_0 of five pitch levels for each subject. As seen in Figure 2, high pitch was produced with lower mean F_0 than mid pitch at the first two time points, and the two pitches had comparable mean F_0 for other time points. Appendix C shows the caregiver errors in the pitch game for all subjects. It was noted that the caregiver chose the mid pitch most frequently among all pitch levels. Therefore it appears that mid pitch was the 'default choice'. Low pitch was produced with the lowest mean F_0 at the first two points, and was similar to rising and falling pitches at the other time points. Given that low pitch attained above-chance accuracy, it is suggested that the beginning of the segment was most significant in perception. It was observed that all target pitches fell at the end; the fall for low, rising and falling pitches were more noticeable than for high and mid pitches.

Subject 2

Subject 2 achieved mid, rising and falling pitches with higher than chance accuracies: all were at 40%. Figure 2 shows that high pitch was produced with the highest mean F_0 at the first four time points. Surprisingly, high pitch attained at-chance accuracy only. The other four pitches had very similar starting points, which was different from Bauer & Benedict (1997)'s findings that mid pitch starts at the mid register and low pitch starts at the low register for normal Cantonese speakers. Low pitch had the lowest mean F_0 at the end, but was between high pitch and mid pitch at time points 2 and 3 (25%, 50%). Although rising pitch attained above-chance accuracy, the acoustic result was surprising. Rising pitch was close to mid pitch from the beginning to time point 4 (75%) as seen in Figure 2. Falling pitch had a distinctive fall in mean F_0 at time point 3 (50%), which was predicted to account for the above-chance perceptual accuracy. It was observed that all target pitches rose at the beginning and fell at the end. From Appendix C, it was noted that the mid and falling pitches were chosen most frequently by the caregiver.

Subject 3

Subject 3 achieved mid and falling pitches with above-chance accuracies. As noted in Figure 2, high pitch was not produced with the highest mean F_0 except at time point 3 (50%) and the end. Although mid pitch attained above-chance accuracy, the acoustic data was surprising. Mid pitch had higher mean F_0 than high pitch at the beginning and time point 4 (75%), and lower mean F_0 than low pitch at time points 2 and 3 (25%, 50%). From Appendix C, it was noted that the mid pitch was selected with the highest frequency by the caregiver. Therefore it appears that mid pitch was the 'default choice'. Low pitch had generally higher mean F_0 than mid pitch and ended with low mean F_0. Falling pitch had a distinctive fall at time point 3 (50%), which was predicted to account for the 60% perceptual accuracy for this target. It was observed that high, mid and rising pitches rose at the end, whereas low and falling pitches fell at the end.

Subject 4

Subject 4 achieved high, mid and rising pitches with higher than chance accuracies: all were at 40%. Figure 2 shows that high and mid pitches had a gentle rise in mean F_0 from the beginning to time point 3 (50%) and fell towards the end. High pitch had the highest mean F_0 at the beginning. Mid pitch had mean F_0 between high and low pitches at the beginning. Given that high and mid pitches attained above-chance perceptual accuracy, it is suggested that the beginning of the segment was most important for perception. Although rising pitch achieved above-chance accuracy, it was surprising that its mean F_0 was similar to low pitch at three time points. It was observed that all pitch targets fell at the end. The standard deviations for all speakers were large for all pitch targets, as shown in Appendix D.

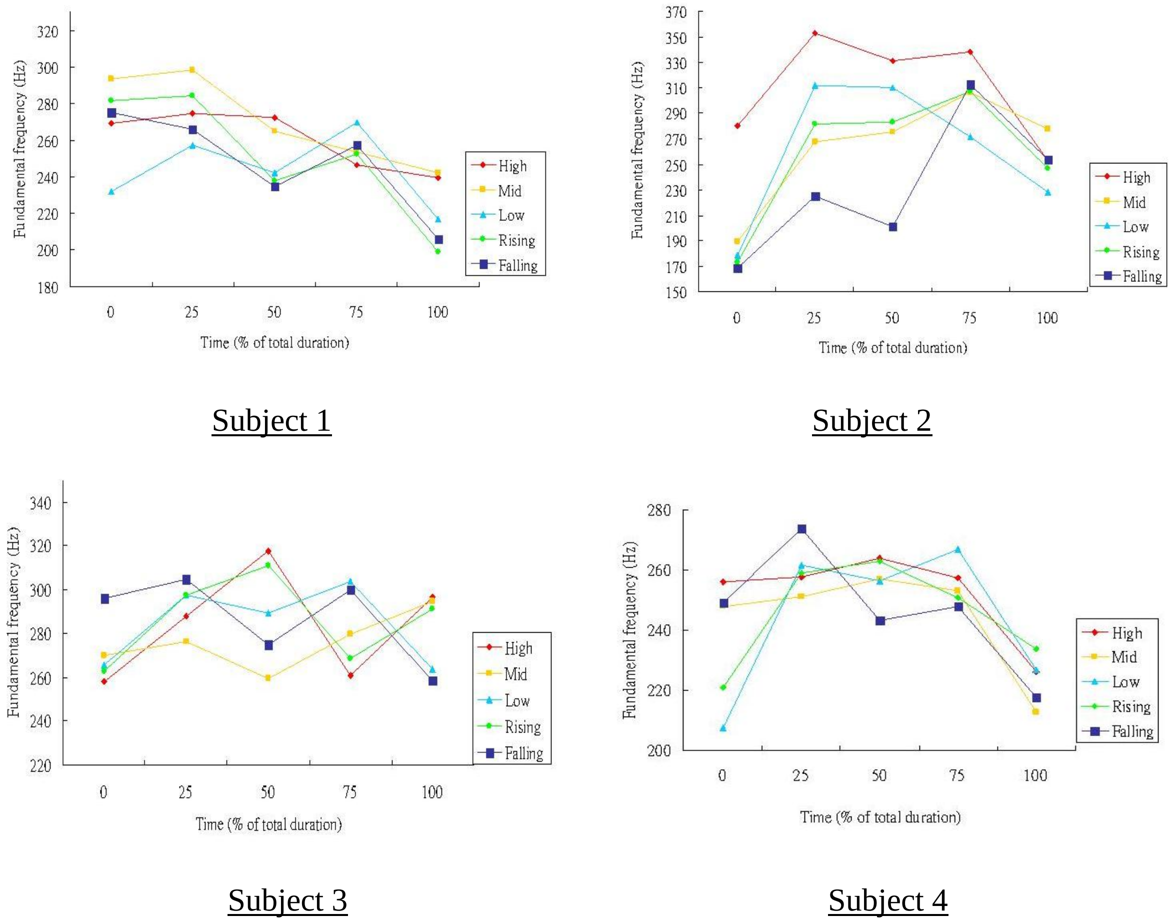


Figure 2. Mean F_0 of five pitch levels for each speaker. Each data point was the average across 5 trials.

Durations

Figure 3 shows that caregiver accuracies at five durations were variable. S1 obtained above-chance accuracy (i.e. above 20%) for only 1 target; S2 for 4 targets; S3 for 2 targets and S4 for all 5 targets. S1 achieved 0% accuracy for the longest duration. S3 attained 0% accuracy for both the long and longest duration. S4 attained 100% accuracy for the longest duration. The shortest duration was identified with 60% accuracy for S2 and S3. The shortest duration was identified with above-chance level for all subjects except S1. The short duration was identified with above-chance accuracy for all subjects except S3.

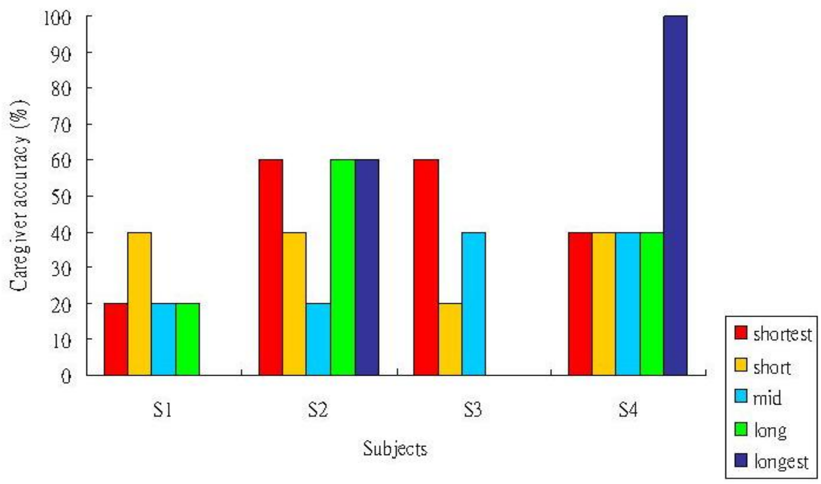


Figure 3. Caregiver accuracy for duration targets.

Figure 4 shows the mean durations and standard deviations for all duration targets using acoustic analysis. As seen in Figure 4, the production of five durations was not as expected for all subjects. The 'longest' duration was produced with the longest mean values for all subjects except S1. The five durations produced by S1 were similar to each other. Notably, the longest mean value was for the 'mid' duration and the shortest mean value was for the 'longest' duration. For S2, the 'shortest' duration had distinctively short mean value. However, the 'short' duration had a longer mean duration than 'mid' and 'long' durations. For S3, the 'shortest' duration was longer than the 'short', 'mid' and 'long' durations (which were similar in durations). The 'longest' target had noticeable long duration value. For S4, the pattern was nearly as expected except for the 'shortest' duration, which had a mean value longer than 'short' and 'mid'. Figure 4 indicates large standard deviations for all subjects. The exact numerical values are shown in Appendix E.

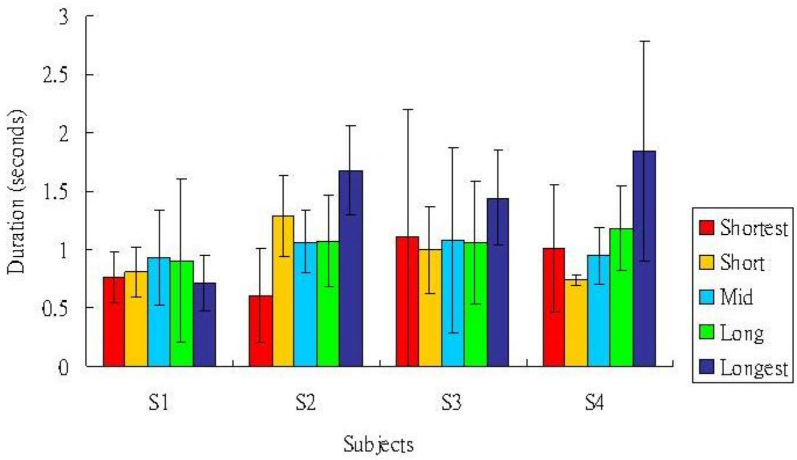


Figure 4. Mean duration and standard deviation of five duration targets. Each mean duration was obtained by averaging across 5 trials. Error bars represent the standard deviation.

The results are now discussed case-by-case.

Subject 1

None of the duration targets attained accuracies above 40%. Only ‘short’ duration achieved above-chance accuracy; ‘shortest’, ‘mid’ and ‘long’ durations were at-chance level; ‘longest’ duration attained 0% accuracy. Appendix F shows the caregiver errors in the duration game for all subjects. It was noted that ‘shortest’ duration was frequently misidentified as ‘short’ duration, which was expected given that mean values of the two categories were close. ‘Short’ duration was frequently misidentified as ‘mid’ duration. This was not surprising because the two categories had similar mean duration values. The errors for ‘mid’ duration were random. ‘Long’ duration was frequently misidentified as ‘mid’ and ‘shortest’ durations. ‘Longest’ duration was frequently misidentified as ‘long’ and ‘shortest’ durations. These errors patterns were not surprising because the measured mean values for the five durations did not differ significantly as seen in Figure 4.

Subject 2

All durations except ‘mid’ duration attained above-chance accuracies; ‘shortest’, ‘long’ and ‘longest’ durations attained 60% accuracies. High perceptual accuracy for ‘shortest’ duration was expected because of its distinctively short mean duration value. As shown in Appendix F, ‘mid’ duration was frequently misidentified as ‘short’ duration. All errors of ‘long’ duration were ‘mid’ duration, which was not surprising because mean durations for the two categories were close. All errors of ‘longest’ duration were ‘long’ duration. This was surprising because mean duration of ‘longest’ duration was distinctively longer than ‘long’ duration as shown in Figure 4.

Subject 3

Only ‘shortest’ and ‘mid’ durations attained above-chance accuracies; both ‘long’ and ‘longest’ durations achieved 0% accuracies. As seen in Appendix F, most errors for ‘short’ duration were ‘shortest’ duration. This was expected because the two durations had similar mean durations as shown in Figure 4. ‘Long’ duration was frequently misidentified as ‘mid’ and ‘short’ durations. This was not surprising because the three categories had similar mean duration values. ‘Longest’ duration was frequently misidentified as ‘mid’ and ‘short’ durations. This was surprising because mean value of the ‘longest’ duration was distinctively longer that all other durations as noted in Figure 4.

<u>Subject 4</u>

All durations attained above-chance accuracies; 'longest' duration attained 100% accuracy. As shown in Appendix F, most errors for 'shortest' duration were 'short' duration. This was surprising because mean duration values of the two categories were not close as shown in Figure 4. Errors for 'short' duration were 'mid' and 'long' durations. 'Mid' duration was frequently misidentified as 'long' duration. 'Long' duration was frequently misidentified as 'longest' duration. This was surprising because mean duration of the 'longest' duration was distinctively longer than the 'long' duration as noted in Figure 4.

DISCUSSION

Pitch levels

The present study found that there was much individual variation in caregiver accuracy in the pitch game. This was consistent with the findings of Patel & Salata (2006). In addition, the dysarthric speakers' acoustic production varied across subjects. Different mean F_0 values across the five time points marked inter-subject differences. Intra-subject differences were noted by large standard deviations at five pitch levels for all speakers. Therefore, the present study supported previous findings that speakers with motor speech disorders demonstrated inter-subject variability and intra-subject variability, which was probably due to their deficit in control of the mechanism for speech production (Weismer & Liss, 1991).

This study was inconsistent with the findings of Patel & Salata (2006) in that the low pitch category was identified most accurately across all subjects in their study. In this study, the mid pitch category was identified with above-chance level for all caregivers. This result has to be interpreted with caution because it was noted that all caregivers except S4 chose mid pitch most frequently among the five pitch levels, which suggested that mid pitch may be a default choice and thus the selection may not reflect what they perceived. Besides, the different pitch levels targeted in two studies might attribute to the discrepancy. Patel & Salata (2006) targeted at three pitch levels (high, mid, low) whereas the current study targeted at five pitch levels (high, mid, low, rising and falling).

The caregivers in this study demonstrated better performance in the pitch game than the caregivers in Patel & Salata (2006)'s study judging from above-chance-level performance. Comparing the number of categories that were identified at above-chance levels, the caregivers in this study attained two to three among five categories while the caregivers in Patel & Salata (2006)'s study attained zero to one among three categories. One explanation is that dysarthric speakers are not homogeneous and that individual differences were expected. It is also possible that the different number of trials (20 trials in Patel & Salata's study versus five trials in the current study) may have contributed to the differences in outcome. The third possibility is that Cantonese-speaking listeners are better than English-speaking listeners in identification of pitch levels because of their experiences with a tonal language. However, this hypothesis would be in contradiction to the finding of Francis & Ciocca (2003) who found that Cantonese and English listeners performed similarly in their experiment. It should be noted that there were several differences between the two studies. Francis & Ciocca (2003) studied lexical tones whereas the current study targeted pitch levels only. Another difference was the number of levels: their study targeted ten levels ranging from high level tone to low level tone while this study targeted five pitch levels (high, mid, low, rising, falling). The third difference was that their listeners participated in a discrimination task while listeners in the present study performed an identification task.

The relationship between acoustic patterns and caregivers' perception of pitch in this study was complex. Acoustic results could be used to explain some of the error patterns in caregiver perception. For example for S1, most errors for high pitch were mid pitch, which could be explained by their overlapping mean F_0 in Figure 2. These findings were consistent with the literature that F_0 is the major cue for pitch perception (Bauer & Benedict, 1997; Fok, 1974). However, acoustic results did not always match with perceptual results. For example for S1, mid pitch was identified with 100% accuracy and was predicted to have a distinctive mean F_0. However, mid pitch was produced with higher mean F_0 than high pitch in 4 of the 5 time points as shown in Figure 2. One account for the poor correspondence between acoustic and perceptual findings is that caregivers used not only acoustic information in perception, but also used cues such as visual cues, voice intensity or voice quality that were not measured in this study. For example, it was observed that S4 tended to raise his upper body when attempting to produce a high pitch. Although it is known that there is a close correspondence between F_0 and pitch perception, this is based on normal speakers. The relationship may be more complex in speakers with dysarthria due to voice disorders, prosodic disturbances, physical movements, variability, etc. The second proposal is related to the nature of the task in this study. The caregivers were asked to make choices which involved chance level, and the perceptual results might not truly reflect what they perceived.

Although the present study found poor correspondence between acoustic and perceptual results for pitch in most circumstances, there were a few occasions that they corresponded to each other. For these cases, it appears that the caregivers attended to the beginning of the segment for perception. This was in contradiction to the findings of Khouw & Ciocca (2007) that F_0 change over the later part (62.5% and 75%) of the vocalic segment was crucial in identifying tone for normal listeners. This discrepancy may be attributed to differences between the two studies. Ten adolescent normal speakers and 30 adult listeners participated in Khouw & Ciocca (2007)'s study, whereas four children dysarthria speakers and four caregivers participated in the current study. In addition, Khouw & Ciocca (2007) studied lexical tone (tones 55, 25, 33, 21, 23, 22), while the current study targeted pitch levels (high, mid, low, rising, falling).

Durations

The present study supported Patel & Salata (2006) that caregiver accuracy was variable across speakers in the duration game. Their finding that the long duration category was the most difficult for caregivers was found for two subjects in the current study, but not the other two. This could be attributed to individual variability in dysarthric speakers, some of whom have better control of respiration and others who have more difficulty in sustaining longer duration because of inadequate respiratory support for speech resulting from reduced vital capacity (Hardy, 1983; Yorkston, Beukelman & Bell, 1988). This study found that both the shortest duration and short duration were identified with above-chance accuracies for all subjects except one. This may reflect that dysarthric speakers generally are more accurate at producing targets of short duration because of impaired respiratory support for speech (Yorkston et al., 1988).

It was noted that caregiver performance in the duration game was generally better in the present study than the study of Patel & Salata (2006). Zero to one among three duration categories was identified at above-chance levels for the five subjects in the previous study, while one to five among five duration categories for the four subjects in the present study. Again, one explanation is that dysarthric speakers are heterogeneous by nature. The number of subjects in both studies prohibits generalizations. The differences are unlikely to be due to differences in the language of the two studies. Although Lai (2002) suggested that duration is another probable clue to show stress in Cantonese besides pitch, duration is also important in signaling stress and intonation in English (Lehiste, 1976; Patel, 2002b) and Fok (1974) stated that syllable duration is not a significant clue to tone perception in Cantonese.

The present study found that the acoustic values of duration usually corresponded to perception of duration. For S1, the mean values of five durations were similar and the expected low perceptual accuracy was confirmed. For S2, the mean values of three durations (shortest, long, longest) appeared distinctive (0.61s, 1.07s, 1.68s), which matched with their high accuracy scores. For S4, the mean value of the longest duration (1.85s) was distinctive from the other categories (1.01s, 0.73s, 0.95s, 1.18s), which was consistent with the 100% accuracy scores found for this target. However for S3, the mean value of the longest duration (1.45s) was noted to be longer than the other four categories (which ranged from 1.06s to 1.10s), but 0% accuracy was found. This is proposed to be related to the differences in duration that allow discrimination. The range of differences in the duration categories that were perceivable in S1, S2 and S4 were from 0.46 to 1.12s, while S3's production ranged from 0.35 to 0.45s. It is suspected that the differences were too small and not distinctive enough for the caregiver to perceive. This hypothesis is supported by Moore (2003) that the detectable durations for duration discrimination were 4, 15, and 60ms for duration of 10, 100, and 1000ms respectively. Hence, it is proposed that there is a cut-off value on differences in duration which allows the caregivers to perceive different categories.

However, there were occasions where acoustic results did not correspond to perceptual results. For example for S2, all errors for the longest duration were the long duration as shown in Appendix F. However, the mean value of the longest duration (1.68s) was distinctively longer than the long duration (1.07s). Therefore, it is hypothesized that although acoustic cue was important for perception of duration, the caregivers also used other cues that were not measured in the present study for perception. For example, the caregiver of S2 reported that she considered the target as a longer duration for trials that the child intended to use more effort in vocalization.

The present study found that all speakers failed to produce five durations with the expected order. This is probably because dysarthric speakers have less flexible respiratory function than normal speakers because of muscle coordination (Hardy, 1983; Yorkston, Beukelman & Bell, 1988), resulting in difficulty in producing vocalizations with the appropriate durations. Although limited range of mean duration among categories were noted in the present study, it was remarkable that dysarthric speakers in this study were able to mark contrasts that were perceivable by listeners within such small range. This has significant clinical implications: dysarthric speakers may be trained to utilize duration as one parameter of input in augmentative and alternative communication devices.

Limitations of the current study

One limitation of the study was the small sample size, which made it difficult to generalize the results to other groups of dysarthric speakers. Another limitation was small number of trials for each target in comparison to Patel & Salata (2006)'s study, which might have lead to the possibility of not representing speakers' abilities to the greatest extent. However, speaker fatigue may confound the results if more trials are requested. The large standard deviations for all speakers for each game should be taken into consideration when considering the mean F_0 and mean duration results.

Clinical implications

Since Cantonese dysarthric speakers in the present study were found to be able to use pitch levels and durations to communicate with their caregivers, the use of prosody as an input to AAC as well as an additional means to communicate with partners is possible. Given that many AAC devices have the capacity to include speech signal, clinicians are encouraged to consider training dysarthric speakers to use these non-verbal vocalizations as input in AAC and to interact with communication partners.

Directions for future research

Future research is suggested to include a larger number of subjects so that results can be more generalized to the dysarthric population. The present study focused on measuring acoustic cues only. Future research might investigate other additional potential cues to the perception of prosody, such as visual cues, voice quality and voice intensity so as to find out cues that are more frequently used and more communicative to caregivers.

To conclude, the present study found that Cantonese-speaking children with severe dysarthria have the ability to use prosody (pitch and duration distinctions) to signal contrasts that were perceivable by their caregivers. This implied that prosody might be used as a means to interact with their caregivers or with augmentative and alternative communication devices.

ACKNOWLEDGMENTS

I would like to thank Professor Tara Whitehill for her guidance throughout the course of dissertation. I would like to thank Dr. Rupal Patel for kindly providing the computer software, willing to facilitate extension in the present study and providing advice during the research process. I would also like to thank Bill Kuo for modification of the computer software, Dr. Valter Ciocca and Ms. Joan Ma for their advice on performing the acoustic analyses. I would also like to thank the speech therapists who kindly helped with subject recruitment. Last but not least, I would like to thank all participants who took part in this research study.

REFERENCES

Ansel, B. M., & Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35*, 296 – 308.

Bauer, R. , & Benedict, P. K. (1997). *Modern Cantonese phonology.* Berlin: Mouton de Gruyter.

Boersma, P. , & Weenink, D. (2000). *PRAAT: A system for doing phonetics by computer (Version 4.0.30) [Computer software].* Amsterdam: Institute of Phonetic Sciences.

Darley, F. L. , Aronson, A. E. , & Brown, J. R. (1975). *Motor speech disorders.* Philadelphia: Saunders.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management.* (2nd ed.). St. Louis: Mosby.

Fok, Y. Y. (1974). *A perceptual study of tones in Cantonese.* Hong Kong: Centre of Asian Studies, University of Hong Kong.

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *Journal of the Acoustical Society of America, 114, 3,* 1611-1621.

Hardy, J. C. (1983). *Cerebral palsy*. Englewood Cliffs, NJ: Prentice-Hall.

Khouw, E. & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*, 104 – 117.

Lai, Y. W. E. (2002). *Prosody and prosodic transfer in foreign language acquisition:*

Cantonese and Japanese. Munchen: Lincom Europa.

Le Dorze, G., Ouellet, L., & Ryalls, J. (1994). Intonation and speech rate in dysarthric speech. *Journal of Communication Disorders, 27,* 1 – 18.

Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225 – 239). New York: Academic Press.

Leuschel, A. ,& Docherty, G. J. (1996). Prosodic assessment of dysarthria: Effects of sampling task and issues of variability. In D. A. Robin, K. M. Yorkston & D. R. Beukelman (Eds.), *Disorders of motor speech: Assessment, treatment and clinical characterization* (pp.155 - 180). Baltimore: Paul H. Brookes Publishing.

Love, R. J. (2000). *Childhood motor speech disability. (2nd ed.).* Boston: Allyn & Bacon.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing (5th ed.).* San Diego: Academic Press.

Patel, R. (2002a). Phonatory control in adults with cerebral palsy and severe dysarthria. *Augmentative and Alternative Communication, 18*, 2 – 10.

Patel, R. (2002b). Prosodic control in severe dysarthria: Preserved ability to mark the question-statement contrast. *Journal of Speech, Language, and Hearing Research, 45*, 5, 858-870.

Patel, R. (2004). The acoustics of contrastive prosody in adults with cerebral palsy. *Journal of Medical Speech Pathology, 12*, 4, 189-193.

Patel, R., & Salata, A. (2006). Using computer games to mediate caregiver-child communication for children with severe dysarthria. *Journal of Medical Speech-Language Pathology, 14, 4,* 279 – 284.

Platt, L., Andrews, G.., & Howie, F. M. (1980). Dysarthria of adult cerebral palsy: II. Phonemic analysis of articulation errors. *Journal of Speech and Hearing Research, 23,* 41 – 55.

Platt, L., Andrews, G., Young, M., & Quinn, P. (1980). Dysarthria of adult cerebral palsy: I. Intelligibility and articulatory impairment. *Journal of Speech and Hearing Research, 23*, 28 – 40.

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica, 33,* 368 – 392.

Vance, J. E. (1994). Prosodic deviation in dysarthria: A case study. *European Journal of Disorders of Communication, 29*, 61-75.

Weismer, G., & Liss, J. M. (1991). Acoustic / perceptual taxonomies of speech production deficits in motor speech disorders. In C. A. Moore, K. M. Yorkston & D. R. Beukelman (Eds.), *Dysarthria and apraxia of speech: Perspectives on management* (pp.245 - 270). Baltimore: Paul H. Brookes.

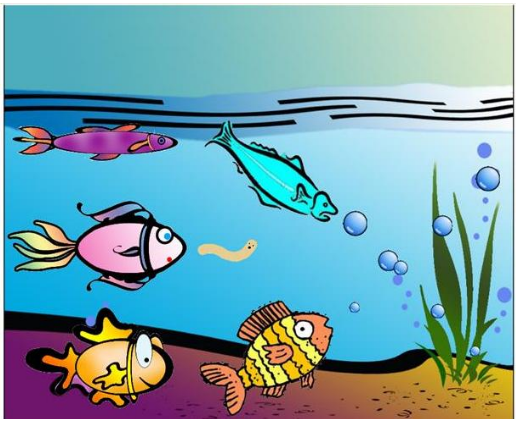
Yorkston, K. M., Beukelman, D. R., & Bell, K. R. (1988). *Clinical management of dysarthric speakers.* London: Taylor & Francis.

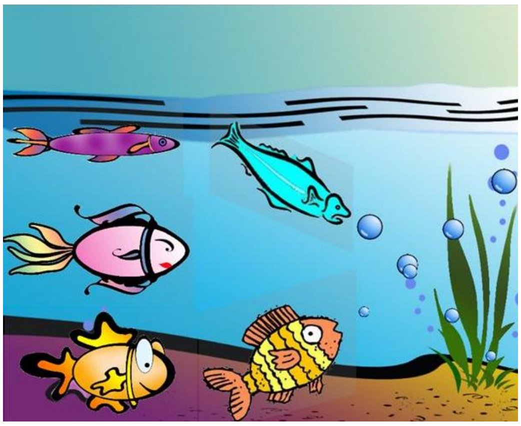
APPENDICES

APPENDIX A

Displays of the child and caregiver interfaces for the pitch and duration games

a) Pitch game





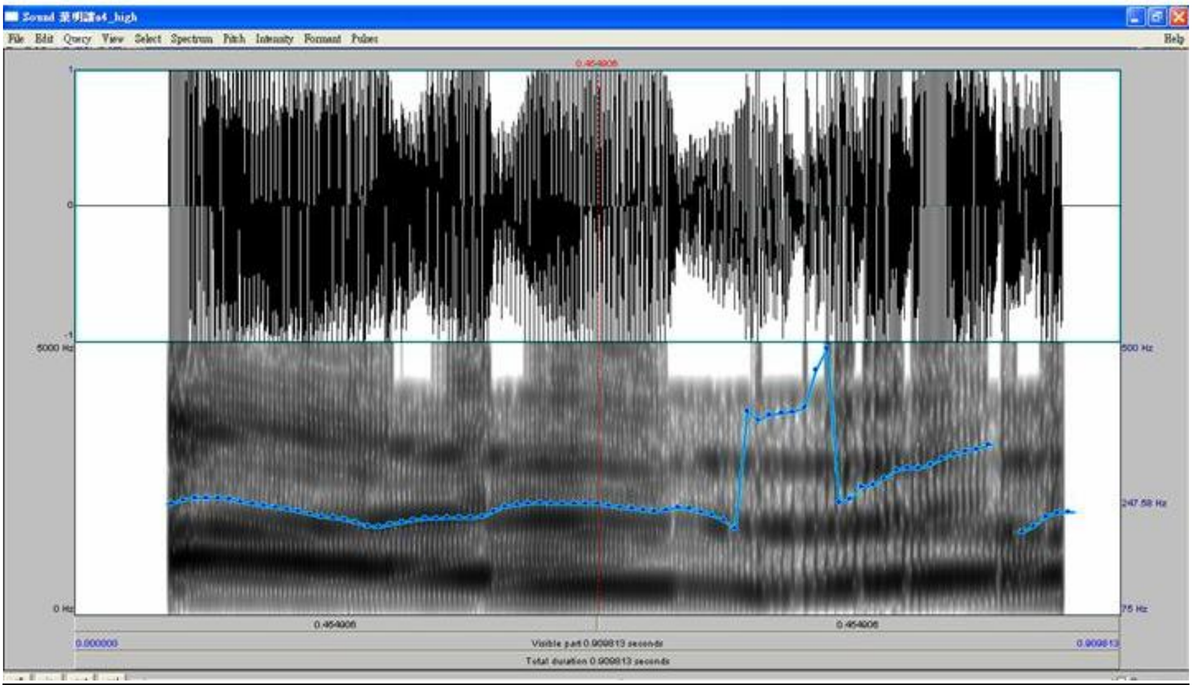
b) Duration game

Child’s interface

Caregiver’s interface

APPENDIX B

Sample of an acoustic signal analyzed using PRAAT



APPENDIX C

Caregiver performance in pitch game

Table C1. Subject 1

Target	Caregiver selection				
	High	Mid	Low	Rising	Falling
High	40% (2/5)	40% (2/5)	---	20% (1/5)	---
Mid	---	100% (5/5)	---	---	---
Low	---	---	40% (2/5)	40% (2/5)	20% (1/5)
Rising	20% (1/5)	40% (2/5)	---	20% (1/5)	20% (1/5)
Falling	40% (2/5)	40% (2/5)	---	---	20% (1/5)

Table C2. Subject 2

Target	Caregiver selection				
	High	Mid	Low	Rising	Falling
High	20% (1/5)	40% (2/5)	---	---	40% (2/5)
Mid	20% (1/5)	40% (2/5)	20% (1/5)	20% (1/5)	---
Low	40% (2/5)	20%(1/5)	20%(1/5)	---	20% (1/5)
Rising	---	20% (1/5)	---	40% (2/5)	40% (2/5)
Falling	---	20% (1/5)	---	40% (2/5)	40% (2/5)

Table C3. Subject 3

Target	Caregiver selection				
	High	Mid	Low	Rising	Falling
High	20% (1/5)	60% (3/5)	---	20% (1/5)	---
Mid	40% (2/5)	40% (2/5)	---	---	20% (1/5)
Low	20% (1/5)	20% (1/5)	20% (1/5)	20% (1/5)	20% (1/5)
Rising	---	40% (2/5)	40% (2/5)	20% (1/5)	---
Falling	20% (1/5)	20% (1/5)	---	---	60% (3/5)

Table C4. Subject 4

Target	Caregiver selection				
	High	Mid	Low	Rising	Falling
High	40% (2/5)	---	---	---	60% (3/5)
Mid	20% (1/5)	40% (2/5)	---	20% (1/5)	20% (1/5)
Low	20% (1/5)	---	20% (1/5)	40% (2/5)	20% (1/5)
Rising	---	20% (1/5)	20% (1/5)	40% (2/5)	20% (1/5)
Falling	---	20% (1/5)	20% (1/5)	40% (2/5)	20% (1/5)

Note. Dashed line represented 0% accuracy. Numbers in parenthesis were the number of correct trials over total number of trials. Highlighted percentages represented accuracies where the targets were selected.

APPENDIX D

Range of standard deviation of the pitch targets

Pitch levels	Subject 1	Subject 2	Subject 3	Subject 4
High	34.10 - 61.00	29.77 - 65.43	18.50 - 72.25	16.57 – 30.06
Mid	12.56 - 59.38	22.44 - 71.47	46.96 - 65.73	10.25 – 21.61
Low	17.54 - 81.02	23.07 - 57.39	53.82 - 79.50	12.90 – 45.00
Rising	16.37 - 44.25	19.93 - 82.14	37.09 - 70.40	11.24 – 57.76
Falling	20.91 - 129.30	10.62 - 84.36	26.34 - 106.76	13.47 – 44.00

Note. The values were in the unit of Hz and corrected to 2 decimal places.

APPENDIX E

Mean and standard deviations of the duration targets

Table E1. Mean durations

			Mean		
Subject	Shortest	Short	Mid	Long	Longest
S1	0.76	0.81	0.92	0.91	0.71
S2	0.61	1.29	1.06	1.07	1.68
S3	1.10	1.00	1.08	1.06	1.45
S4	1.01	0.73	0.95	1.18	1.85

Note. The values were in the unit of second and corrected to 2 decimal places.

Table E2. Standard deviations of duration targets

			Standard deviation		
Subject	Shortest	Short	Mid	Long	Longest
S1	0.22	0.21	0.40	0.70	0.24
S2	0.40	0.35	0.27	0.39	0.38
S3	1.10	0.37	0.80	0.52	0.41
S4	0.55	0.05	0.24	0.36	0.94

Note. The values were in the unit of second and corrected to 2 decimal places.

APPENDIX F

Caregiver performance in duration game

Table F1. Subject 1

Target	Caregiver selection				
	Shortest	Short	Mid	Long	Longest
Shortest	20% (1/5)	40% (2/5)	20% (1/5)	---	20% (1/5)
Short	---	40% (2/5)	40% (2/5)	20% (1/5)	---
Mid	20% (1/5)	20% (1/5)	20% (1/5)	20% (1/5)	20% (1/5)
Long	40% (2/5)	---	40% (2/5)	20% (1/5)	---
Longest	40% (2/5)	---	20% (1/5)	40% (2/5)	---

Table F2. Subject 2

Target	Caregiver selection				
	Shortest	Short	Mid	Long	Longest
Shortest	60% (3/5)	20% (1/5)	---	---	20% (1/5)
Short	---	40% (2/5)	20%(1/5)	20% (1/5)	20% (1/5)
Mid	---	60% (3/5)	20% (1/5)	20% (1/5)	---
Long	---	---	40% (2/5)	60% (3/5)	---
Longest	---	---	---	40% (2/5)	60% (3/5)

Table F3. Subject 3

Target	Caregiver selection				
	Shortest	Short	Mid	Long	Longest
Shortest	60% (3/5)	20% (1/5)	20% (1/5)	---	---
Short	60% (3/5)	20% (1/5)	---	---	20% (1/5)
Mid	---	20% (1/5)	40% (2/5)	20% (1/5)	20% (1/5)
Long	20% (1/5)	40% (2/5)	40% (2/5)	---	---
Longest	20% (1/5)	40% (2/5)	40% (2/5)	---	---

Table F4. Subject 4

Target	Caregiver selection				
	Shortest	Short	Mid	Long	Longest
Shortest	40% (2/5)	40% (2/5)	---	20% (1/5)	---
Short	---	40% (2/5)	20% (1/5)	40% (2/5)	---
Mid	---	---	40% (2/5)	40% (2/5)	20% (1/5)
Long	---	---	20% (1/5)	40% (2/5)	40% (2/5)
Longest	---	---	---	---	100% (5/5)

Note. Dashed line represented 0% accuracy. Numbers in parenthesis were the number of correct trials over total number of trials. Highlighted percentages represented accuracies where the targets were selected.